DETECTING DDoS ATTACK USING HYBRID MACHINE LEARNING ALGORITHMS

Yesubai Rubavathy[1], Ezhil Arasan[2], Jaya Prakash Anthony[3], Maharajan[4]

[1]Assisstant Professor/CSE,
[2,3,4]UG scholars, CSE Department, Francis Xavier Engineering College

---

***Abstract*:** With the increase in usage of networking technology *and the Internet, Intrusion detection becomes important and challenging security problem. A number of techniques came into existence to detect the intrusions on the basis of machine learning and deep learning procedures. This paper will give inspiration to the use of ML and DL systems to IP traffic and gives a concise depiction of every one of the ML and DL strategies. This paper gives an audit of 40 noteworthy works that covers the period from 2015 to 2019. ML and DL methods are compared with regard to their accuracy and detection potential to detect different types of intrusions. Future Research includes ML and DL methods to find the intrusions so as to improve the detection rate, accuracy and to minimize the false positive rate.*

Keywords— Machine learning, Deep Learning, intrusions, attacks, network security, datasets, metrics.

I. INTRODUCTION

Security has become a necessity with the development of the Internet and networking technology. It is a set of rules and configurations considered to protect network from various security attacks. Most of the security mechanisms must provide the following security services [1].

A. Confidentiality

Protecting data from unauthorized users.

B. Availability

Ensures the data is available to legitimate users at all times

C. Integrity

Ensures that data cannot be altered or duplicated or replayed during transmission

D. Non-Repudiation

Ensures that user does not refuse that he has used the network.

E. Access Control

Controlling the access of unauthorized use of resources.

F. Authentication

Authentication means that the identity of the user is certain

Intrusion is an action that negotiates the three security requirements: Confidentiality, Availability and Integrity of the resources. Intrusion Detection system is an application that inspects network systems for any intrusions. Any suspicious activity will be reported to the administrator or centrally controlled security information and event management system. Whenever a suspicious activity is detected, it issues alerts. IDS's are used to detect various security attacks. [1].

Security Attacks classification [2]

Attack is an indication of security compromise. Security attacks are classified as:

A. Viruses

A virus is a self-copying program that affects and spreads through files. Usually it affixes itself to the files, which will cause it to be run when the file is launched. There are several types of viruses such as, Macro Viruses, System and Boot Record Infectors, File Infectors.

B. Worms

These are self-reproducible programs that spread through the network. They do not need an infected file to circulate. Worms are of two types: mass-mailing worms and network-aware worms.

C. Trojans

A Trojan gives off an impression of being evil, yet for the most part have some dangerous reason. Trojans hold some payload, for example, viruses, information obliteration and remote access strategies.

D. Logic Bombs

Logic bomb is a different form of Trojan that releases its payload when some condition is met.

E. Buffer overflows

Buffer overflows make use of wrong programming methods in which buffers are allowed to be overloaded. The data filling in the buffer which is filled beyond its capacity can overflow into the neighboring memory, then it can either control information or it is utilized to change the program execution. Buffer overflows are of two types of. Heap Buffer Overflows, Stack Buffer Overflow

F. Denial of Service Attacks

These attacks usually interrupt the network service or a system, so that it cannot be used further and to degrade the performance.

Three types of DoS attacks are there: distributed, network based and host based.

G. Network-Based Attacks

These attacks affect the networks that operate the protocols. These are Spoofing (IP Spoofing and MAC Address Spoofing), Wireless attacks, Session Hijacking and Web application attacks.

H. Wireless Network Attacks

The majorities of the wireless network attacks is due to false configurations and generally require MAC address spoofing to obtain full control. These are Wired Equivalent Protocol (WEP) cracking.

I. Web Application Attacks

These are the attacks that targets web applications. Basically the application layer in the TCP/IP protocol stack will be attacked. There are different ways a web application can be attacked. Hidden Field Manipulation, Cross Site Scripting, Database Attacks, Cookie Poisoning, Parameter Tampering.

J. Information Gathering Attacks

In this the intruder simply gets important information or unauthorized access to data without having to launch an attack. Information gathering is passive. Various possible attacks are sniffing, scanning or probing and mapping.

K. Password Attacks

An attacker wishes to obtain control over the user account or control of a computer, will use a this attack to get the required password. Various possible attacks are Exploiting the Implementation, Dictionary Attack or Password Guessing, Brute Force [2].

Classification of IDS [3]

The intrusion detection system is categorized on factors: Location, Detection Mechanism. . In view of the position or location, IDS's are categorized as Host based IDS, Network based IDS. . In view of the detection mechanism IDS's are classified as Misuse detection, Anomaly detection.

A. NIDS

NIDS examine the network traffic and analyze it for any attacks. The IDS is set along the network boundary or between the network and the server. The benefit of this framework is that it very well may be conveyed effectively and requiring little to no effort, without being loaded for every system.

B. HIDS

HIDS is a single computer specific IDS which monitors the security of the system from internal and external attacks. The IDS is installed on the system itself. Favorable position of this framework is it can precisely screen the entire framework and doesn't require establishment of some other equipment.

C. Anomaly Detection

It depends on the theory that the normal user behavior differ from attacker behavior with respect to bandwidth, protocol, ports and other devices. Anomaly based IDS's are used to detect novel attacks. Example, Bro-IDS, Snort [4].

D. Misuse or Signature based IDS

It compares attack patterns with those that are saved in the database to detect unusual behavior. It detects only known attacks. Example, Suricata [5] is Signature based IDS.

Challenges of anIDS:

- IDSs are infamous for creating false positives.
- IDS can't be a one-size-fits all setup to work precisely and viably. Also, this requires an insightful IDS examiner to tailor the IDS for the interests and needs of a given site. Learned prepared framework experts are rare.
- The hoax with IDS is that you need to realize what the attack is to have the option to distinguish it.

An IDS has the capacity to detect attacks over the huge network without human intervention. The techniques based on deep learning, soft computing and machine learning are used to detect intrusions.

II. MACHINE LEARNING (ML)

People are using machine learning techniques since 1969. Arthur Samuel characterized ML as a field of study that enables machines to learn without being explicitly modified [6].The ML Techniques consists of two stages. Training or Learning and Testing or Inference [7].

A. Learning phase

First, the machine learns by exploring patterns. The list of features used to solve the task is known as a feature vector. A feature vector is used to resolve the problem and it is a subset of dataset. The machine uses some sort of algorithms to convert reality to a model. Hence the learning stage is used to wrap up data to a model [7].

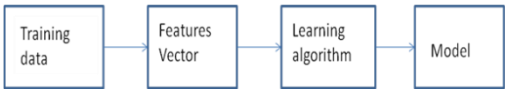

Fig 1. Learning Phase

B. Inference phase

Once the model is created, we have to test this model using new data. This new data will be converted to a feature vector; it is applied to a model and gives a prediction. There is no need to change the rules or educate again the model. We can use the earlier model to make conclusion on the new data [7].

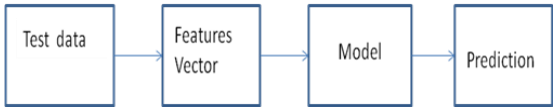

Fig 2. Inference Phase

Application of ML in IP traffic classification [8]:

We can apply ML concepts to IP traffic classification. We define the subsequent three terms identifying with flows or streams:

A. Flow or Uni-directional stream

An arrangement of bundles utilizing a similar five-tuple: source and goal IP ports, convention number and source and goal IP addresses.

B. Bi-directional stream

It is a mix of two uni-streams heading in inverse di between a similar source and target IP locations and ports.

C. Full-stream:

A bi-directional stream caught in its full lifetime, from start as far as possible of the association.

Various packets fit in the same flow constitute the instance. The attributes used for the packets for different flows are the features. All the features are not important. We have to eliminate the noisy features, redundant and irrelevant features.

Dimensionality reduction or Feature reduction

Reducing amount of features of the dataset is Feature Reduction. Reduction of the features can be done using feature extraction and feature selection.

A. Feature Selection Algorithms

Feature set quality is important for the performance evaluation of ML algorithms. Redundant or irrelevant features cause negative impacts on the performance of ML algorithms. It makes the system costlier.

Feature selection process is followed by four basic steps [9]

1. Subset Generation**:** It is a process of generating the candidate feature subsets for evaluation phase, using various search strategies (Best first, Random search. Greedy search etc.)
2. Subset Evaluation**:** Numbers of subsets generated in the previous phase are evaluated by comparing it with the previous best one based on some evaluation criteria. The better turned out subset will replace the previously chosen subset.
3. Stopping Criteria: The above two processes continues repeatedly until some stopping criteria is not reached.
4. Result validation**:** The best selected subset is validated by using some kind of test e.g. implementing the classifiers algorithm or some clustering technique etc.

Feature Selection Algorithms

Algorithms for feature selection can be categorized into wrapper method, filter method, embedded or hybrid method [10].

1. Filter methods

These methods use statistical tests for the selection of features for the correlation with target value. Based on their scores in the tests features are selected [10]. This method comes under univariate feature selection (looks at each feature independently of the others)

- Pearson's Correlation [9]
- Chi-Square [11]
- Signal to noise ratio
- F-score
- Information Gain [12]
- Analysis of variance(ANOVA)

Fig 3. Filter method processing

2. Wrapper methods

Wrapper methods use ML algorithms that will finally be used for learning the model that means its results influence toward the Machine Learning algorithm used. Based on the conclusions drawn from the earlier model, we choose either to add or delete features from our subset [11]. These methods comes under multivariate feature selection (considers all features simultaneously).

Forward feature selection

In this method we are not considering any features in the beginning. It is a repetitive method in which we keep adding the features to the model in each rerun till accumulation of new features does not influence the performance.

Backward feature elimination

In this method, we can start with all the features and delete the insignificant features at each step. This process is repeated this process until convergence.

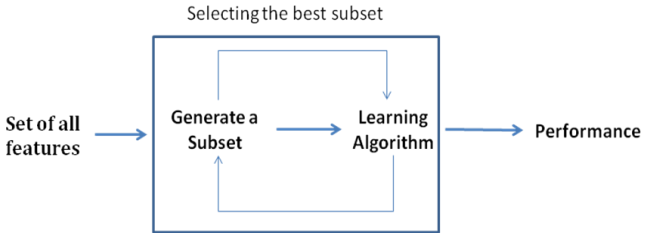

Fig 4. Wrapper method processing

Recursive feature elimination

It constantly makes the models and keeps the best or worst features at each rerun. It develops the following models with the left over features. This procedure is rehashed until every one of the features are finished. At that point it positions the features dependent on the request for their disposal.

3. Hybrid or Embedded methods

These techniques consolidate the characteristics of filter and wrapper strategies. It is executed by algorithms that have their pre-defined feature selection strategies.Data analysis such as regression or classification can be done in the reduced features more accurately than in the original ones.

B. Feature extraction techniques [13]

These methods do the transformation of data from high level to low level dimensional space. Examples include

- Generalized Discriminant Analysis (GDA)
- Principal Component Analysis (PCA) [14]
- Linear Discriminant Analysis (LDA)

ML Techniques:

After Feature reduction, ML algorithms are applied to train the system and to classify the data into normal or attack data. There are four approaches of ML techniques. They are

1. Supervised learning

It is analogous to the supervision of a teacher to the class. This algorithm uses labeled data for

training and predicts the new

data based on training it gained. They are classified as classification and regression. Various techniques fall under this category are:

- Bayesian network [15]
- Gaussian process regression
- Lazy learning
- Decision trees [16][17]
- K-Nearest Neighbor Algorithm [18]
- Information fuzzy networks[19][20]
- Support vector machines [11]
- Bootstrap aggregating
- Linear regression
- Naive Bayes classifier [21]
- Hidden Markov models [22]
- Artificial Neural Network[23]
- Random Forests [24]

2. Unsupervised learning

Unsupervised learning uses unlabeled data and finds structures, patterns, knowledge in the data.[5] They are classified as clustering and association. Various .algorithms fall under this category are:

- Information bottleneck method
- Vector Quantization
- Expectation-maximization algorithm
- Generative topographic map
- Self-organizing map
- Apriori algorithm
- K-means algorithm [25]
- Single-linkage clustering
- Fuzzy clustering [26]
- Local Outlier Factor

3. Semi-supervised learning

A combination of both labeled and unlabelled data[6]. Algorithms under this category are:

- Co-training
- Low-density separation
- Generative models
- Graph-based methods

4. Reinforcement learning

This method often learns from the environment repetitively to take actions that would maximize the reward or minimize the risk. Algorithms under this category are:

- Q-Learning[27][28]
- Temporal Difference (TD)[29]
- Markov decision process
- Deep Adversarial Networks[30] [31]
- Monte Carlo methods

Brief Description of Algorithms:

1. Decision Trees

Decision tree is used to approximate discrete valued and continuous valued attributes. Decision trees resemble if-then-else conditions. There are three fundamental components in a decision tree: decision node, branch and leaf node. Classification starts at the root node, testing the attribute determined by that node, at that point descending the tree limb comparing to the estimation of the attribute in the given example. For testing the attributes at a particular node we use different techniques like information gain, entropy, chi-square, gini etc. The major decision tree algorithms are CART, C4.5 [16], ID3 [17], LMT Tree, etc. Decision tree algorithm can handle binary or multiclass classification problems.

2. Naive Bayes Classifier [32]

It is based on the Bayes theorem. Bayes theorem calculates posterior probability as follows.

$$p(h/D) = p(D/h)\,p(h)\,/P(D) \quad (1)$$

where P(h|D) is the posterior probability of the hypothesis h. P(h) is the prior probability of hypothesis h. P(D|h) denotes the probability of observing data D given some world in which hypothesis holds. P(D) is the prior probability of training data.

Naive Bayes Classifier:

$$v_{NB} = argmax_{v_j \epsilon V} P(v_j) \prod_k P(a_i|v_j) \quad (2)$$

Method**:**

- Transform the dataset into frequency table
- Prepare likelihood table
- Use Bayesian equation to find the posterior probability

For a given test instance, Class label with largest posterior is picked as the expected class [32] [33]. There are three Naive Bayes algorithms: Bernoulli Naive Bayes [21], Multinomial Naive Bayes [34], and Gaussian Naive Bayes [35]. It can handle multiclass classifications under the assumption of conditional independence

3. Support Vector Machine (SVM)[36]

In supervised method of SVM, we plot the data items of the dataset as points in the n-dimensional space where n is the number of attributes. Each feature value can be plotted as a coordinate in n-dimensional space. Then we find some hyper plane that splits data into different groups such that closest points from each group to the hyper plane are the farthest from the hyper plane. SVM can handle binary classification but can be extended to handle multiclass with additional constraints and parameters added to the optimization problem. For supervised learning algorithm Multi-class SVM is used. An unsupervised learning algorithm, One class SVM is useful for novelty detection [37].

4. Artificial Neural Network (ANN)

The goal of the ANN is to solve the problems in the same way as the human brain does. This technique depends on natural neural systems structures and standards [38]. Artificial neurons are utilized to construct ANN [39]. Weights related with the neurons are refreshed over the learning procedure of neural systems [40]. The ANN comprises of hidden layer, input layer, output layer. The structure of the ANN can be isolated into two classes: feed forward and recurrent networks. In feed-forward systems, inputs are sustained a single way to the outputs.

In recurrent systems, among inputs and outputs there are forward and in reverse ways. ANN can deal with multiclass classifications.

5. Genetic Algorithm:

These are the evolutionary algorithms, based on Darvin's Theory, the principle of survival of the fittest [41]. There are four functions used in this process. They are selection, initialization, mutation and crossover. The algorithm works by repeatedly updating population, a pool of hypotheses. In each iteration, fitness function is utilized to evaluate the members of the population. Then a best population is selected using the selection function as an optimal solution and combines these selected individuals to create the next generation population using crossover function. Crossover and mutation operations are applied to generate new offspring [42].

6. Hidden Markov Model:

This is a statistical model and is an alternate of finite state machine. HMM has a set of prior or initial probabilities, hidden states, transition probabilities, an output alphabet or observations, output or emission probabilities. This model determines the hidden observations from the output observations. For that it uses a forward - backward correlation [42]. HMM model computes the joint probability of hidden states given the probability of observed states [43]. We determine the best the sequence with the highest probability and choose that sequence as the best sequence of hidden states.

7. Swarm Intelligence

A swarm is a collection of collaborative and cooperating agents that work jointly to resolve a problem and provide the best possible solution [44]. The main principle behind swarm intelligence system is that it can act in a synchronized manner exclusive of the external controller or a coordinator. On the whole the system behavior results from the interactions of the neighbors and its surroundings. They are classified into: Ant Colony Clustering (ACC) based IDS, Particle Swarm Optimization (PSM), Ant Colony Optimization (ACO) based IDS, Shuffled Frog Leaping Algorithm, Fruit Fly Optimization Algorithm (FOA), Firefly Algorithm (FA), Flower Pollination Algorithm (FPA), Cuckoo Search Algorithm (CS), Bat Algorithm (BA), Wolf Pack Algorithm (WPA) etc.

8. Fuzzy Logic

Dual logic's truth values, either totally false (0) or totally true (1) are used for reasoning, but these kinds of limitations are not there in Fuzzy logic [45]. That means in Fuzzy logic the statement's truth value range is in between 0 and 1, inclusive of 0 and 1.

9. K-Means

It is an unsupervised algorithm used to solve clustering problem. It considers k-clusters. Data points in clusters are heterogeneous to peer groups and data points inside clusters are homogeneous.

Method:

- k-means picks k-centroids, one for each group and allocate every datum point to the bunch with least separation between its centroid and information point.
- With the existing cluster data points recalculate the centroid. Reassign each data point to the nearest group centroid.
- Repeat the above step until convergence occurs that means centroids will be the same.

10. kNN (K-Nearest neighbors)

This algorithm uses full dataset as training data and stores that into memory. Each time a new instance is to be classified, it finds k-instances from the training data that are comparative or closest to the given new instance. The nearest neighbors of the instance or similarity between the instances are calculated using one of the Manhattan distance, Euclidean distance, Minkowski, and Hamming distance.

Single classifiers:

Individual classifiers follow different objectives to develop a single classification mode. Single classifiers create and evaluate a single learning algorithm. A given algorithm may perform better than all others for a specific subset of problems. No algorithm achieves the best accuracy for all situations.

Multiple classifiers or Ensemble of Classifiers

Combining predictions given by multiple classifiers for classification is known as ensemble learning. Compared to individual base learners, the ensemble of classifiers provides a stronger generalization capability [46] We can use different names for multiple classifiers: ensemble methods, combination of classifiers, classifier fusion, classifier aggregation etc. We can combine predictions of classifiers using voting or non-voting methods. In voting counts of each classifier are used to classify the new example. Vote of each classifier is weighted by its performance measure on the training data. In non-voting Class probabilities of all models are summed up by specific rule.(sum, product, max, min, median etc). Multiple classifiers can take decisions either group wise or by specialized or dynamic integration(select only those classifiers which are more accurate for the new examples).Multiple classifiers can be homogeneous or heterogeneous. Homogeneous classifiers use the same algorithm on different datasets. Heterogeneous algorithms use different algorithms over the same data. The methods of ensemble learning are: bagging, boosting, voting, stacking etc [47].

III. DEEP LEARNING (DL)

DL [48][49] is a subset of ML that imitates the communication of neurons in a brain, because it makes use of deep connections of neurons, It is called as deep learning. In Neural Network architecture the layers are placed one above the other. To learn from the data, the machine uses different layers. The strength of the model is characterized by the number of layers in the model. Characteristics of the DL algorithms are:

a. DL finds the features needed for classification automatically. So feature extraction is not needed in DL for each problem.

b. DL systems need large amounts of data for learning. They would not perform well when there are small volumes of data.

c. Learning models engineered under different learning frameworks differs.

d. DL algorithms take longer training times and smaller testing times.

e. For implementation, DL algorithms require high-end machines with GPU's.

f. It is difficult to interpret and understand DL algorithms because of the complex network structures.

g. DL algorithms can also be divided into supervised, unsupervised and semi-supervised.

Deep learning Techniques:

- Deep Neural Network (DNN)[50]
- Deep Belief Networks[51][52]
- Deep Auto encoders[53]
- Restricted Boltzmann Machines[54]
- Hopfield Neural network
- DBNs or RBMs or Deep Auto encoders Coupled with Classification Layers
- Recurrent Neural Networks[55]
- Self-Organized Map (SOM)
- Bi-directional Neural Networks (Bi-ANN)[56]
- Feed-forward neural networks (FFNN)[56]
- Convolutional Neural Networks [57]
- Generative Adversarial Networks[58]
- Long – Short - Term Memory (LSTM)[59]

1. Deep Neural Network (DNN):

DNN is built of an , a number of hidden layers, input layer and an output layer. Back propagation algorithm is used for training them and to lessen the error between the target value and the actual value.

2. Feed-forward neural networks (FFNN):

It is a DNN in which information is fed into the network only in a single direction that is in forward direction. The information that is fed into the network can be transformed into an output. Supervised learning is used for training FFNN.

3. Recurrent Neural Networks (RNN):

Recurrent networks are a type of ANN's that apply to time series data. ANNs have edges fed into the next layer at the same time step but RNNs have edges fed into the next time step. RNNs use feedback loops which are connected to the previous data. RNNs have two input sources, the present and the previous; both are combined to determine the new instance. RNNs have memory.

4. Deep Auto encoders:

These are a type of neural networks that contains the same input and output. They compress the input into a hidden format and then rebuild the output from this hidden format. An Auto-encoder has an encoder, a decoder. Their combination comprises a deep network. We can create a new feature set that has lesser dimensions using an encoder than that of its input for the hidden levels. From the learned representation, the decoder function rebuilds its input.

5. Restricted Boltzmann Machines (RBM):

RBMs are undirected, two-layer, bipartite, graphical models that form the building blocks of DBNs. RBMs are unlabelled and typically prepared each layer in turn. The main layer is the input layer; the subsequent layer is the hidden layer. It has no connections within the same layer, but, there can be inter layer connections between input and hidden layers.

6. Deep Belief Networks (DBN):

DBN is a store of limited Boltzmann machines, in which each RBM layer talks with both the past and coming about layers. The centers of any single layer don't talk with each other. Unsupervised training is used for RBM layers. Supervised training is used for last layer which is fully connected. Each hidden layer is individually trained to rebuild the inputs by adjusting weights.

7. Long-short term Memory (LSTM):

To update the weights of the network, Neural Networks use Back Propagation (BP) algorithm. BP first calculates the gradients from the error using the chain rule and then it updates the weights. Updating the weights in deep neural networks might face some problems. As we go back with the gradients, it is possible that the values get either smaller exponentially which causes Vanishing Gradient problem or larger exponentially which causes Exploding Gradient problem. Due to this we get the problems of training the network. RNN's neglects to confine long term reliance that connects consecutive tasks and the base of inclination drop which is disappearing. The essential objective of LSTM is to accomplish vanishing gradient descent which is an enhancement calculation to discover artificial neural systems loads to stay away from long haul reliance issues [60].LSTM uses four neural networks and various memory blocks called cells. Memory manipulations in the cells are done by gates. LSTM uses three gates: output gate, input gate, forget gate. At any step, inputs to the LSTM cell are previous memory state, current input, previous hidden state. Outputs from the LSTM cell are current memory state and current hidden state. Components of LSTM are: Output gate(NN with sigmoid), Forget gate (NN with sigmoid), Input gate(NN with sigmoid), Candidate layer(NN with Tanh), Hidden State(vector), Memory state(vector).

IV. NETWORK SECURITY DATASETS

In the next few subsections, we are providing overview of various datasets used for network intrusions over the years. The basic requirements for any security research are reasonable use of data and correct choice of data. The volume of the dataset also influences the training models of the Machine Learning and Deep Learning methods. The information can be gained in two different ways: straightforwardly as direct access and utilizing a current open dataset. Direct access means utilizing programming tools to acquire new packets.

Existing datasets can save the information gathering time and improves the research productivity [61]. The most common datasets used for NIDS that are publicly available include DARPA, NSL-KDD[62], KDD CUP 99[63], UNSW-NB[64], ADFA-WD[65], ADFA, ADFA-LD[65], CICIDS2017[66], CIDDS-001[67] dataset etc.Most commonly used HIDS datasets are KDD Cup 98[68], UNM[69], KDDCup 99. These datasets excludes recent attacks of networks. G. Creech and J. Hu [70] planned new datasets called ADFA-WD, ADFA-LD and made them to be publicly available.

A.UNSW-NB dataset:

The research team of cyber security ACCS (Australian Centre for Cyber Security) launched a dataset called UNSW-NB15 which is used to resolve the issues found in the KDD Cup 99[71]. For this dataset, the attacks are classified as follows:

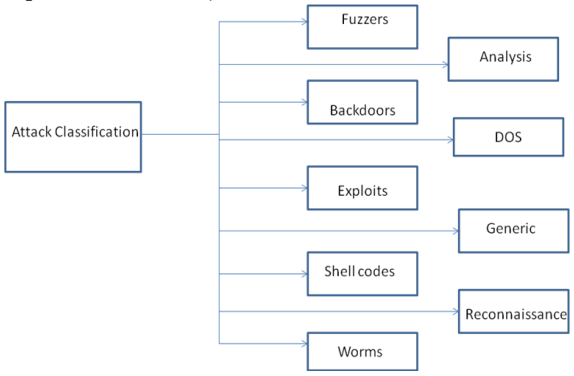

Fig 5. Attacks in UNSW-NB dataset

B.KDD CUP 99 dataset:

KDD Cup 1999 dataset was created by MIT Lincon laboratory. The dataset contains 5 classes (DoS, Normal, U2R, R2L, Probe) [71]. The features of this dataset are grouped into different classes as shown in the figure below.

- Probe or scanning: Intruder attempts to obtain data about the target machine.
- Denial of Service (dos): prevent authentic clients from accessing a service.
- User to Root (u2r): Intruder has nearby access to the machine and attempts to gain privileges of a root machine.
- Remote to Local (r2l): Intruder does not have access to a machine, hence attempts to obtain access.

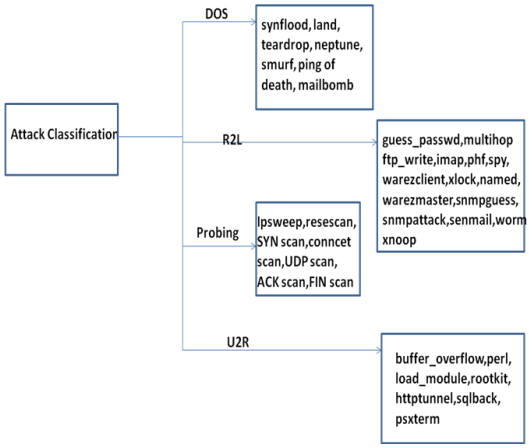

Fig 6 Attacks in KDD CUP 99 dataset

C.CICIDS 2017 Dataset:

The dataset contains the behavior of benign and attacks that shows the on going systems traffic. Various attacks are classified as follows [104]:

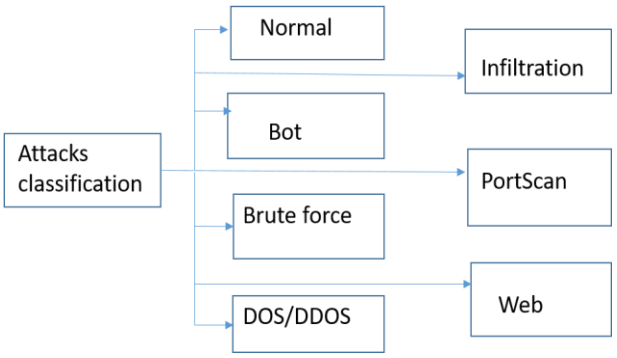

Fig 7. Attacks in CICIDS 2017 dataset

D.ADFA Windows(ADFA-WD) / ADFA Linux (ADFA-LD)

ADFA Linux is a series of system calls dataset that was gathered in networks of current operating systems. The ADFA Windows consists of dynamic link libraries (DLL) and system calls for various attacks. [44]. The information of the ADFA-LD and ADFA-WD dataset is given below.

Dataset	ADFA-LD		ADFA-WD	
	Traces	System calls	Traces	System calls
Train	833	308077	355	13504419
Validation	4372	2122085	1827	117918735
Attack	746	317388	5542	74202804
Total	5951	2747550	7724	205625958

Fig 8. ADFA-LD / ADFA-WD data

E. CIDDS-001 dataset [71]:

This dataset comprises of 13 features. The attacks are categorized into five classes as follows.

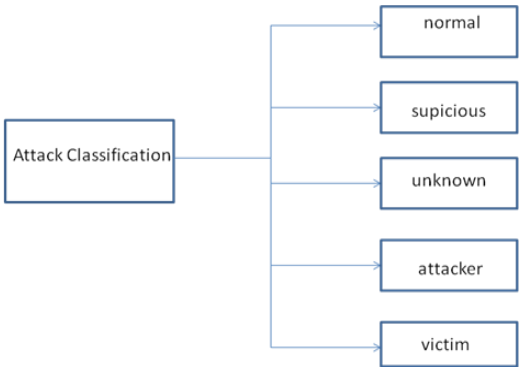

Fig 9. Attacks in CIDDS-001 dataset

V PERFORMANCE EVALUATION METRICS

The confusion matrix is an array that portrays the classification results showing whether they are effectively classified or not. Confusion matrix is a 2x2 matrix for a binary classification. For a binary classification, the results are alienated into four categories as shown in Table 1 as follows.

Table-I: Confusion Matrix

	Predicted as Positive	Predicted as Negative
Labeled as Positive	True Positive(TP)	False Negative(FN)
Labeled as Negative	False Positive(FP)	True Negative(TN)

- *True Positive* (TP): attack data correctly classified by the model as an attack. It indicates when an alarm is generated and there is an intrusion.
- *False Negative* (FN): attack data that is wrongly classified by the model as normal data. It indicates when an alarm is not generated but there is an intrusion.
- *False Positive* (FP): normal data that is wrongly classified by the model as an attack. It indicates when an alarm is generated but there is no intrusion i.e., a false alarm in this case.
- *True Negative* (TN): the samples that are correctly classified by the model as negative. It indicates when an alarm is not generated and there is no intrusion. It is a correct rejection.

Further, the following metrics can be calculated from the confusion matrix:

- *Accuracy*: accuracy = (TP + TN)/(TP + TN + FP + FN). Ratio of the number of correctly classified samples to the total number of samples for a given test data set.
- *Positive Predictive Value(PPV) or Precision*: Precision = TP/(TP + FP). It calculates the ratio of all correctly detected items to all actually detected items.
- *Sensitivity or Recall or True Positive Rate (TPR) or Probability of Detection (PD) or Detection Rate(DR)*: DR = TP/(TP + FN). It calculates the ratio of all correctly detected items to all items that should be detected.
- *False Negative Rate (FNR): FNR = FN/(TP + FN)*. The ratio of the number of is classified positive samples to the number of positive samples.
- *False Positive Rate (FPR) or False Alarm rate (FAR) or Fall-out* : FAR = FP/(FP+TN) = 1-TNR. The ratio of the number of misclassified negative samples to the total number of negative samples.
- *specificity or True Negative Rate (TNR): TNR* = TN/(TN + FP). The ratio of the number of correctly classified negative samples to the number of negative samples.
- *F1-score*: It calculates the harmonic mean of the precision and the recall. F1-score = 2 * TP / (2 * TP + FN + FP) or F1-score = 2*precision*recall / (precision + recall).
- *ROC*: In A receiver operating characteristic (ROC) space, the abscissa for each point is FPR and the ordinate is TPR, which also describes the trade-off of the classifier between TP and FP. ROC's main analysis tool is a curve drawn in ROC space - the ROC curve. A higher the value for the area under the ROC curve indicates that the classifier is better at distinguishing between the classes[1].
- *AUC*: The value of AUC is the size of the area under the ROC curve. In general, AUC values range from 0.5 to 1.0, and larger AUCs represent better performance.

Fig 12. Classification Metrics

According to the performance analysis of different ML methods, the results are analyzed, compared and validated using NSL KDD and KDD CUP 99 because both the datasets have same set of features.

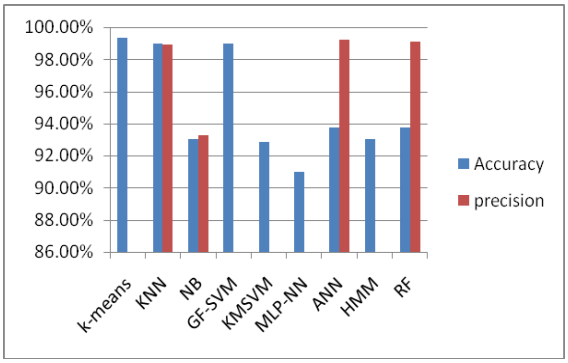

Fig 10. Performance comparison of different ML

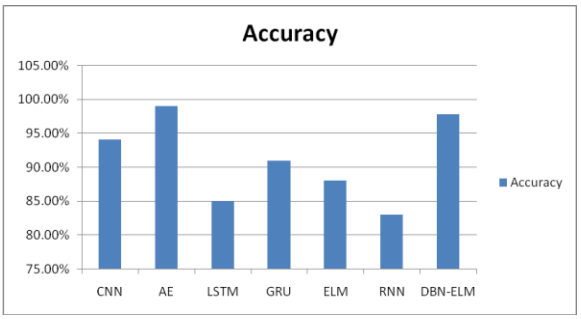

Fig 11. Performance comparison of different DL classifiers.

VI. LITERATURE REVIEW

Sumaiya Thaseen Ikram et al. in [11] projected a model with multi class support vector machine (SVM) and chi-square feature selection. Over fitting constant and Radial Basis Function kernel parameter are used for optimization in parameter tuning. They have used NSL-KDD dataset for evaluating the model. They have achieved 98% accuracy and 0.13 as false alarm rate for the implemented model. Alex Shenfielda et al. in [23] proposed a new model suitable for use in deep packet inspection based intrusion detection systems using ANN architecture. The proposed ANN architecture achieved an accuracy of 98%, an AUC of 0.98in repetitive Tenfold cross validation. An online exploit and vulnerability repository exploitdb [103] was used to classify the attacks. Nabila Farnaaz and M. A. Jabbar in [24] employed an IDS approach with random forest classifier. Feature selection method used is symmetrical uncertainty of attributes. NSL-KDD data set was used for evaluating the new approach.They have compared random forest modeling with j48 classier. Their results proved that MCC, accuracy and DR used for multiclass attacks classification are increased. They have achieved 0.99% of MCC, 0.00527% of FAR, 99.84% ofDR and 99.67% of AC. Rishabh Das and Thomas H. Morris in [42] proposed a model using the four classifiers: Random Forest, Naive Bayes, OneR, J48. They have used MODBUS data collected from a gas pipeline for evaluation. Tenfold cross validation they have used. Algorithm J48 performed better than others with Area Under Curve of 0.995, Precisionof 0.992, Recall of 0.992. Mostafa darkaie, Reza Tavoli in

[56] employed a novel approach using multi-layer perceptron (MLP) neural network. Back propagation algorithm is used to train the neural network and minimize the error associated with weights. In this model, they have used KDD99 dataset for training and evaluation. In this model, for feature reduction, PCA algorithm is used. The new method has achieved 91% accuracy..Sara A. Althubiti in [60] proposed anomaly based IDS using Long Short Term Memory (LSTM). CIDDS dataset is used for evaluating the model. They have achieved an accuracy of 85.5%.R.Vinaykumar in [71] proposed a model called scale-hybrid-IDS-AlertNet, a hybrid framework of NIDS and HIDS that can be used to detect and to alert possible cyber attacks.

The benchmark datasets used are : CICIDS 2017, NSLKDD, Kyoto, UNSW-NB15, WSN-DS and also KDD CUP99.

Classifiers:

Traditional classifiers.Mehrnaz Mazinia et al. in [72] proposeda

method for an anomaly based IDS (A-NIDS) using fusion of artificial bee colony (ABC) and AdaBoost algorithms. ABC algorithm is used for feature selection and AdaBoost is used to evaluate and classify the features. The proposed method is tested on NSL-KDD and ISCXIDS2012 datasets. AdaBoost has shown DR of 99.61%, 0.01%FPR, 98.90% of AC with ABC feature selection method. In [74] Jabez J, Dr. B. Muthukumar employed a novel approach named as outlier detection. In this, Neighborhood Outlier Factor (NOF) is used to measure the anomaly dataset. The model is trained with big datasets with distributed storage environment for evaluating the performance of IDS. The proposed model takes less execution time, high detection rate than other approaches. Ripon Patgiri et al. in [75] developed a model that using Random Forest and Support Vector Machine. Recursive feature Elimination is used as a feature selection method with SVM. NSL KDD dataset is used for training and evaluation. They have used Cross-validation to evaluate it. Using their model, Random Forest performed better than SVM before feature selection. After feature selection, SVM performed better than RF.Simone A. Ludwig in [76] developed a neural network ensemble method for classification of attacks. The ensemble method consists of an extreme learning machine, an auto encoder, a deep neural network and a deep belief neural network. NSL-KDD data set is used for evaluation. They have obtained a precision of 93%, f-measure of 92% and recall of 92% . Saroj Kr. Biswas in [77] proposed an IDS model that compares performance of different combinations. A subset of significant features is selected using feature selection algorithms and then to train the model different classifiers are used. NSL-KDD dataset is used to evaluate the model using 5-fold cross validation. CFS, IGR, PCA, and minimum redundancy maximum- relevance feature selection techniques, and SVM, DT, k-NN, NB, and NN classifiers are used in their model. Through their observations it is proved that the performance of K-NN is superior to other classifiers and performance if IGR is superior to other feature selection methods. Mohammad Almseidin et al. [78] were performed several experiments to estimate the performance of the ML classifiers: Random Tree, J48, MLP, Random Forest, Naive Bayes, Decision Table, and Bayes Network. They have used KDD intrusion detection dataset for testing. Among these RF classifier obtained an accuracy of 93.77% - highest accuracy than others and with the smallest false positive rate and RMSE value. Chuanlong Yin et al. [79] employed a model using recurrent neural networks (RNN-IDS). They have compared it with those of traditional ML classifiers. NSL KDD dataset is used for training and evaluation of the model. Their experimental outcome proved that performance of RNN-IDS is better than that of traditional classification methods. OLASEHINDE Olayemi O et al. in [80] have built Naïve Bayes, KNN and Decision tree intrusion detection Models with a consistency features selection reduced training dataset, the models were evaluated using the testing dataset, from the work. The results of their evaluation on the UNSW-NB15 dataset show that; Decision Tree has the highest overall model classification accuracy of 86.77%.

In [81] PEIYING TAO et al. projected a model for IDS using Genetic Algorithm and Support vector machine i.e., FWP-SVM-GA (feature selection, weight, and parameter optimization of support vector machine based on the genetic algorithm). In their Paper, a two-step optimization of SVM based on GA is used : to select the feature subset (FWP-SVM-GA-1) and to optimize the selected feature weights and SVM parameters (FWP-SVM-GA-2). KDD CUP 99 dataset is used for evaluation. Experimental results showed a raise in accuracy, detection rate, true positive rate and drop off in SVM training time and false positive rate. In [82] YIHAN XIAO et al. proposed a network intrusion detection model based on a convolutional neural network(CNN_IDS). PCA and AE are used for dimensionality reduction. To evaluate the model, KDD-CUP99 dataset was used. The experimental results showed that accuracy, detection rate, and false alarm rate reaches 94.0%, 93.0%, and 0.5% respectively. HONGYU YANG AND FENGYAN WANG in [83] proposed a model using improved convolution neural network (IBWNIDM). Training the model consists of a forward and backward propagation process. They have applied NSL-KDD CUP data set for evaluation. They have compared the model with other three methods - typical neural networks, Recurrent Neural Network and Deep Belief Nets. They have achieved an accuracy of 95.36%, higher than the other three models.In [84] Prof. Ujwala Ravale proposed hybrid technique (KMSVM) that combines RBF kernel function of SVM and K-Means clustering for classification. The proposed technique is evaluated using KDD-CUP99 data Set. K- Means clustering is used for feature reduction. The experimental results showed that the performance of the proposed model is superior to others. Accuracy of the proposed method is 92.86%.Parisa Lotfallahtabrizi, Yasser Morgan in [85] presented Host Intrusion Detection System (HIDS) that uses Artificial Neural Networks (ANNs). FFNN (feed forward NN) with two hidden layers are used. They have used smart-devices data for evaluation. They got values as FAR is 0.011 and DR is 0.974.In [86], Zhida Li et al. proposed a model using two Recurrent neural networks (Long Short Term Memory, Gated Recurrent Unit) and Broad Learning System (BLS) and its extensions. They have evaluated performance of proposed models using BGP datasets (Border Gateway Protocol datasets) and also the NLS-KDD dataset. With BGP dataset, they have achieved f-score and accuracy in the range of 90%-95% and with NSL-KDD dataset these are in the range 80%-85%. Feng Chen et al. [87] proposed a model with K- Nearest Neighbors. In that model, they have used a tree-seed algorithm (TSA) forfeature reduction and KNN is used for classification. KDD CUP 99 datasets and UCI repository datasets are used to evaluate the proposed model. They have compared the proposed model with genetic algorithm and basic particle swarm optimization (PSO) combined with KNN.

With KDD CUP dataset the model has achieved an accuracy of 87.34%. Hafsa BENADDI in [88] proposed PCA-Fuzzy Clustering-KNN method. Principle component analysis is used for feature reduction. For training and evaluation, NSL KDD Dataset was used. Results showed that detection effectiveness of the proposed algorithm is superior than other algorithms. Roshan Pokhrel in

[89] proposed a model using

Naïve Bayes (NB) and Support Vector Machine (SVM). For analyzing the model they have used data from two Organizations. The proposed model achieved an precision of 95%-96%, recall of 98-99% and accuracy of 92%-93%, and ROC curve area is 0.9313 and 0.9518 respectively for two Organizations.

In [90], Bayan Alsughayyir proposed Deep learning approach - Deep Auto encoders for multi-class classification. Min-Max Scaler is used for normalization process. NSL KDD dataset was used for training and evaluation. The performance of the proposed approach proved as superior to the traditional methods with an accuracy of 91.28% for the testing phase. Azar Abid Salih in [91] proposed a model with three classifiers: Naive Bayes, Multilayer Perceptron and K-Nearest Neighbors. For Feature selection three methods are used: Gain Ratio, Information Gain, Correlation. The proposed model is analyzed using KDD-CUP 99 data set. The KNN got the highest detection rate than others. Li Yong, Zhang Bo in [92] proposed an algorithm based on Convolutional neural network. Batch Normalization algorithm and Inception model are added to this algorithm to improvise the convergence speed of the model. This algorithm is verified by KDD-Cup 99 data. Experiment results proved that the model has higher accuracy of 94.11% and detection rate of 93.21% than classical classification algorithms. In [93], Hossein Gharaee et.al. implemented an anomaly based IDS using Genetic algorithm and Least Squares Support Vector Machine (LSSVM) named as GF-SVM. Genetic algorithm is used for feature reduction. To solve the local optima problem LSSVM is used. The proposed model is evaluated on UNSW-NB15 and KDD CUP 99 datasets. The proposed has model achieved an accuracy of 99%. D.Selvamani in [94] proposed a model uses different techniques for feature selection: Symmetrical Uncertainty, Chi-Square analysis, Gain Ratio, Information Gain with different Classification methods: Naive Bayes, Support Vector Machine, Artificial Neural Network,. They have used KDD CUP 99 dataset for comparing and evaluating different algorithms. With ANN, NB, SVM classifications, highest accuracy is achieved with information gain feature selection method. Nathan Shone in [95] proposed stacked nonsymmetric deep auto encoder (NDAE) with Random Forest for classification and it is evaluated using the datasets - KDD Cup 99 and NSL-KDD**.** NDAE is used for dimensionality reduction. The result was 99.59%. Arijit Chandra in [96] proposed a hybrid model using Sequential Minimal Optimization (SMO) and K-Means Clustering techniques for classification. They have used Filter-based methods for feature reduction. KDD99 dataset is used for training and testing the proposed model. They have achieved an accuracy of 99.32%. HONGYU YANG et al. in [97], proposed a model for deep belief networks that consists of multilayer with Restricted Boltzmann Machine structure, Back propagation network layer, SVM classification layer. The model was evaluated using NSL KDD dataset. They have achieved a recall of 97%, an accuracy of 97%, precision of 97% and F1 score of 98%.

XIANWEI GAO et al. in [98] proposed an ensemble learning model-Multi tree and adaptive voting algorithm. They have used NSLKDD dataset for training and evaluation and achieved an accuracy of the adaptive voting algorithm as 85.2% and that of the Multi Tree algorithm as 84.2%.

PENG WEI et al. in [99] proposed a new joint optimization algorithm based on Deep Belief Network. To optimize the multi hidden layer NN in the DBN model, they have used particle swarm optimization (PSO), genetic algorithm optimization PSO (GA-PSO) algorithm and an artificial fish swarm algorithm optimization PSO (AFSA-PSO) algorithm. They have used NSLKDD dataset for testing and achieved an accuracy of 83.86%. Liang Dai, Peisheng Pan in [100] proposed DBN-ELM method that uses deep belief network to train NSLKDD dataset and fed them back to the extreme learning machine for classification. They got false alarm rate of 1.81% and an accuracy of 97.82%. Sulaiman Alhaidari and Mohamed Zohdy in [101] proposed a new model using 2D Hidden Markov Model based Viterbi algorithm. They have achieved an accuracy of 92%.Sandeep Ankush Maske in [102] introduced a method called ELM (Extreme Learning Machine) approach. They have used KDDCup99, UNM datasets and ADFA Linux dataset. They have achieved detection

VIII. CONCLUSION

A review has been made of the significant works in the fields of machine learning and deep learning that are used to detect NIDS and or HIDS during the period of 2015 to 2019. All these woks have emphasized distinct machine learning and deep learning strategies utilized, informational index or datasets utilized, assessment measurements for every one of the systems utilized. To decide the powerful approach a few criteria must be considered. The criterion included classification time, training time, detection rate and accuracy. It is hard to distinguish a superior methodology of ML and DL strategies dependent on just one factor like accuracy. On the off chance that the ML and DL strategies looked at dependent on precision or accuracy, these techniques ought to be prepared on

VII. COMPARISION OF DIFFERENT METHODS

Table - II. Comparison of different ML and DL techniques

Machine Learning Methods:

Reference	year	Proposed Model	Dataset used	Feature reduction algorithm used	Attacks detected	Performance	Problem domain
Alex Shenfielda [23]	2018	ANN	Exploit exploitdb	-	Malicious shell code files. Binary classification	Accuracy - 98% Precision - 97% Sensitivity - 95%	NIDS
Arijit Chandra [96]	2019	K-means clustering SMO	KDD CUP 99	Cfs subset Best first	Multiclass classification	Accuracy-99.3262%	NIDS
Azar Abid Salih[91]	2019	KNN, NB, MLP	KDD CUP 99	Information gain, gain ratio, correlation	Binary classification	Accuracy: KNN- 99% MLP-97% NB-93% Precision: KNN-98.9% MLP-96.5% NB-93.3%	NIDS
Jabez J[74]	2015	NOF Neighborhood Outlier Factor	big datasets with distributed storage environment	-	Binary classification	-	NIDS
Nabila Farnaaz [24]	2016	RF	NSL KDD	symmetrical uncertainty	Multiclass classification (DOS,probe,R2L ,U2R)	AC-99.67% DR-99.84% FAR-.0052 MCC- 0.99	NIDS
Mehrnaz Mazinia[72]	2018	Ada Boost	NSL-KDD ISCXIDS2012	Ant bee colony	Multiclass classification	DR- 99.61%, FPR-0.01 AC-98.90%	Anomaly based NIDS
Sumaiya Thaseen Ikram[11]	2017	SVM	NSL-KDD	Chi-square	Multiclass classification	AC-98% FAR-0.13	NIDS
Ripon Patgiri[75]	2018	SVM RF	NSL-KDD	Recursive feature Elimination with RF, SVM	Multiclass classification	-	NIDS
XIANWEI GAO[98]	2019	Multi Tree (Classification Regression tree), Adaptive voting algorithm	NSL-KDD	-	Multi class classification	MT-AC-84.23% AV-AC-85.2%	NIDS

Saroj Kr. Biswas [77]	2018	k-NN, DT, NN, SVM , NB	NSL-KDD	CFS, IGR, PCA, minimum redundancy maximum-relevance	Multiclass classification	KNN-better performance	NIDS
Mohammad Almseidin [78]	2018	J48, RF, MLP, NB, Random Tree, BayesNetwork DecisionTable	KDD	-	Multiclass classification	RF- highest accuracy rate 93.77% Precision -99.1%	NIDS
Rishab Das[42]	2017	NB, RF, OneR, J48.	MODBUS data	-	Multiclass classification	J48-AUC-0.995 J48-Precision- 0.992 J48-Recall-0.992.	NIDS
OLASEHINDE Olayemi O[80]	2018	NB KNN DT	UNSW-NB15	-	-	DT-AC-86.77%	NIDS
PEIYING TAO [81]	2018	FWP-SVM-GA	KDD CUP 99	GA	Binary classification	FWP-SVM-GA1 DR – 96.61% FPR – 3.39 FNR-0.07 FWP-SVM-GA2 DR - 100% FPR -0 FNR-0.07	NIDS
Ujwala Ravale[84]	2015	KMSVM	KDD CUP 99	K-means	Multiclass classification	AC-92.86%	NIDS
Parisa Lotfallahtabrizi [85]	2018	ANN with BP	Online smart-device data	-	Binary classification (normal, attack)	FAR 0.011 DR 0.974	Signature based HIDS
Feng Chen[87]	2018	TSA-KNN	UCI repository and KDDCUP99	Tree Seed Algorithm	Binary classification	KDDCUP: Accuracy-87.34%	NIDS
Hafsa Benaddi[88]	2018	PCA-FC-KNN	NSL-KDD	PCA	Binary Multiclass	K=4 given max DR. DR is 0.53s	Anomaly based NIDS
Roshan Pokhrel[89]	2019	Hybrid model-SVM, NB(semi supervised)	Custom data collected from Two Organizations	-	Binary classification	accuracy of 92%-93%, precision of 95%-96%, recall of 98-99% and ROC curve area is 0.9313 and 0.9518	HIDS
D.Selvamani [94]	2019	ANN NB SVM	KDD CUP	IG GR SU ChiSquare	Binary classification	Accuracy with IG: With NN:93.76% With NB-71.24% With SVM-75.33%	NIDS
Hossein Gharaee [93]	2016	GF-SVM	KDD CUP 99 UNSW-NB 15	GA	Multiclass classification	KDD CUP: AC-99%	Anomaly based NIDS
Sulaiman Alhaidari [101]	2019	2D HMM based viterbi algorithm	NSL-KDD	-	Binary classification	AC-93%	NIDS
Mostafa darkaie , Reza Tavoli[56]	2019	MLP-NN with BP(hybrid model)	KDD CUP 99	PCA	Binary classification	AC-91%	NIDS
Deep Learning Methods							

Reference	year	Proposed Model	Dataset used	Feature reduction algorithm used	Attacks detected	Performance	Problem domain
Li Yong, Zhang Bo[92]	2019	Multi-scale CNN	KDD CUP 99	Batch normalization	Binary classification	AC-94.11% DR-93.21%	NIDS
Bayan Alsughayyir [90]	2019	Deep Auto Encoders	NSL-KDD	Min-max-scaler	Multi class classification	Accuracy-99% F1 score-0.99	NIDS
Nathan Shone[95]	2018	Stacked NDAE with RF	NSL-KDD KDD CUP 99	NDAE	Multiclass classification	Accuracy-99.59%	NIDS
R Vinaykumar [71]	2019	DNN	KDD CUP 99	-	Multiclass classification	-	NIDS, HIDS (hybrid model)
Zhida Li[86]	2019	LSTM GRU BLS	BGP NSL-KDD	-	Binary classification	BGP: LSTM-AC-92% GRU-AC-91% KDD: LSTM-AC-82.78% GRU-AC-82.87% BLS-AC-82.2%	Anomaly based NIDS
HONGYU YANG [97]	2019	DBN-SVM	NSL-KDD	One hot encoding	Multi class classification	AC - 97% Precision - 97% recall - 97% F1 score - 98%.	NIDS
YIHAN XIAO [82]	2019	CNN-IDS	KDD CUP 99	PCA, AE	Multiclass classification	AC- 94.0% DR -93.0% FAR - 0.5%	
HONGYU YANG[83]	2019	IBWNIDM	NSL-KDD CUP	-	Multiclass classification	AC-95.36% TPR-95.55% FPR-0.76%	NIDS
PENG WEI[99]	2019	AFSA-GA-PSO-DBN	NSL-KDD	-	Multi class classification	AC-82.36%	NIDS
Liang Dai, Peisheng Pan[100]	2019	Improved DBN-ELM	NSL-KDD	DBN	Multi class classification	AC-97.82% FAR-1.81	NIDS
Sandeep Ankush Maske[102]	2016	ELM	ADFA Linux dataset.	-	Multi class classification	AC-88% FAR-2.1	HIDS
Sara A. Althubiti[60]	2018	LSTM	CIDDS-001	-	Multiclass classification	AC-85%	anomaly based NIDS
Simone A. Ludwig [76]	2019	ensemble method - AE,DBF, deep NN, and ELM	NSL-KDD	-	Multiclass classification	precision - 92% recall - 93% fscore - 92%	NIDS
Chuanlong Yin[79]	2017	RNN-IDS	NSL-KDD	-	Binary, Multiclass	DR-83.28%	NIDS

same training information and tried on same exact testing information. Therefore, the results are not comparable. Model needs to be retrained long-term and quickly. Future research requires ML and DL strategies to discover the intrusions in order to improve the accuracy, precision and discovery rate and to limit the false positive rate with focus on lifelong learning.

REFERENCES

1. William Stallings, "Cryptography and Network Security", 5th edition, Pearson Education
2. Simon Hansman. "A Taxonomy of Network and Computer Attack Methodologies" 2003, url: https://www.cosc.canterbury.ac.nz/research/reports/HonsReps/2003/hon s_030.pdf (visited on 11/09/2019).
3. Abhishek Pharate , Harsha Bhat , Vaibhav Shilimkar "Classification of Intrusion Detection System" IJCS Volume 118 – No. 7, May 2015.
4. SNORT. (2017). Snort 2.9.7.6. [Online]. Available:https://www.snort.org/
5. OISF. (2018). Suricata 4.0.4. [Online]. Available:https://suricataids.org/about/
6. T. Sree Kala, Dr .A. Christy "A Survey and Analysis of Machine Learning Algorithms for Intrusion Detection System" Jour of Adv Research in Dynamical & Control Systems, 04-Special Issue, June 2017.
7. https://www.guru99.com/machine-learning-tutorial.html
8. Thuy T.T. Nguyen and Grenville Armitage "A Survey of Techniques for Internet Traffic Classification using Machine

Learning" IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL. 10, NO. 4, FOURTH QUARTER 2008.

9. Shilpa Bahl and Dr. Deepak Dahiya "Features Contribution for Detecting Attacks of an Intrusion Detection System" Global Journal of Pure and Applied Mathematics. ISSN 0973-1768 Volume 13, Number 9 (2017), pp. 5635-5653
10. 10.https://www.analyticsvidhya.com/blog/2016/12/introduction-to-feature-selection-methods-with-an-example-or-how-to-select-the-right-variables/
11. Sumaiya Thaseen Ikram , Aswani Kumar Cherukuri "Intrusion detection model using fusion of chi-square feature selection and multi class SVM" Computer and Information Sciences (2017) 29, 462–472.
12. H.Güneş Kayacık, A. Nur Zincir-Heywood, Malcolm I. Heywood "Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets" Conference Paper · January 2005.
13. https://www.geeksforgeeks.org/dimensionality-reduction/
14. W. Wang and R. Battiti, "Identifying intrusions in computer networks with principal component analysis," in Proc. IEEE 1st Int. Conf. Availability Rel. Security (ARES), 2006, p. 8.
15. Wojciech Tylman "Anomaly-Based Intrusion Detection Using Bayesian Networks" Third International Conference on Dependability of Computer Systems DepCoS-RELCOMEX 2008.
16. J. R. Quinlan, C4.5: Programs for Machine Learning. San Francisco, CA, USA: Morgan Kaufmann, 1993.
17. V. H. Garcia, R. Monroy, and M. Quintana, "Web attack detection using ID3," in Professional Practice in Artificial Intelligence. Cham, Switzerland: Springer Int., 2006, pp. 323–332
18. LONGJIE LI, YANG YU, SHENSHEN BAI, YING HOU, XIAOYUN CHEN "An Effective Two-Step Intrusion Detection Approach Based on Binary Classification and k-NN" IEEE ACCESS VOLUME 6, 2018.
19. K.S. Anil Kumar and Dr. V. NandaMohan, " Novel Anomaly Intrusion Detection Using Neuro-Fuzzy Inference System ", IJCSNS International Journal 6 of Computer Science and Network Security, vol.8, no.8, pp.6-11 , August 2008
20. Shingo Mabu, Nannan Lu, Kaoru Shimada,Kotaro Hirasawa, "An intrusion-Detection Model Based on Fuzzy Class-Association-Rule Mining Using Genetic Network Programming", IEEE Transactions On Systems, Man, And Cybernetics Part C: Applications And Reviews, VOL. 41, NO. 1, PP: 130-139 , 2011